

ABSTRACT

A stopword detection component detects stopwords (also stop-phrases) in search queries input to keyword-based information retrieval systems. Potential stopwords are initially identified by comparing the terms in the search query to a list of known stopwords. Context data is then retrieved based on the search query and the identified stopwords. In one implementation, the context data includes documents retrieved from a document index. In another implementation, the context data includes categories relevant to the search query. Sets of retrieved context data are compared to one another to determine if they are substantially similar. If the sets of context data are substantially similar, this fact may be used to infer that the removal of the potential stopword(s) is not material to the search. If the sets of context data are not substantially similar, the potential stopword can be considered material to the search and should not be removed from the query.